Detecting Abrupt Changes in the Spectra of High-Energy Astrophysical Sources

Raymond K. W. Wong[*1], Vinay L. Kashyap[†2], Thomas C. M. Lee[‡3], and David A. van Dyk[§4]

[1]Department of Statistics, Iowa State University
[2]Harvard-Smithsonian Center for Astrophysics
[3]Department of Statistics, University of California, Davis
[4]Statistics Section, Department of Mathematics, Imperial College London

December 10, 2015

Abstract

Variable-intensity astronomical sources are the result of complex and often extreme physical processes. Abrupt changes in source intensity are typically accompanied by equally sudden spectral shifts, i.e., sudden changes in the wavelength distribution of the emission. This article develops a method for modeling photon counts collected from observation of such sources. We embed change points into a marked Poisson process, where photon wavelengths are regarded as marks and both the Poisson intensity parameter and the distribution of the marks are allowed to change. To the best of our knowledge this is the first effort to embed change points into a *marked* Poisson process. Between the change points, the spectrum is modeled non-parametrically using a mixture of a smooth radial basis expansion and a number of local deviations from the smooth term representing spectral emission lines. Because the model is over parameterized we employ an ℓ_1 penalty. The tuning parameter in the penalty and the number of change points are determined via the minimum description length principle. Our method is validated via a series of simulation studies and its practical utility is illustrated in the analysis of the ultra-fast rotating yellow giant star known as FK Com.

[*]Raymond K. W. Wong is an Assistant Professor in the Statistics Department at Iowa State University. Email: raywong@iastate.edu

[†]Vinay L. Kashyap is an Astrophysicist at the Harvard-Smithsonian Center for Astrophysics. Email: vkashyap@cfa.harvard.edu

[‡]Thomas C. M. Lee is a Professor in the Statistics Department at the University of California, Davis. Email: tcmlee@ucdavis.edu

[§]David A. van Dyk is a Professor in the Statistics Section at Imperial College London. Email: dvandyk@imperial.ac.uk

1

1   Introduction

Astronomical sources that exhibit temporal variability or periodicity are among the most scientifically rich in the observed Universe. Variation in source intensity can result from rotations, eclipses, magnetic activity cycles, or the turbulent flows of matter into deep gravitational wells. Massive stellar explosions known as *super novae*, for example, appear as abrupt peaks in the electromagnetic radiation emitted from a source. They sometime produce a narrow beam of intense high-energy radiation that appears as γ-ray bursts from Earth—although there are other sources of γ-ray bursts. These bursts originate in distant galaxies, can last from milliseconds to minutes, and are among the most energetic events that occur in the Universe. There are many less dramatic objects that produce temporal variability; they include signals with repeated peaks and troughs that may or may not follow a predictable periodic pattern. A *variable star*, for example, exhibits fluctuating emission that can be due to a natural expansion and contraction in its radius as it evolves (a so-called *pulsating variable star*) or to eruptions in its coronae such as flares or mass ejections. In extreme cases, superflares can erupt producing millions of times more energy than a typical solar flare; if such a flare occured on the Sun it could destroy the Earth's ozone layer and cause mass planetary extinction. Although the mechanism is not well understood, super flares are most likely to occur on fast rotating stars (McKee, 2012).

In this article we develop statistical methods that enable us to identify changes in observed emission from astronomical sources, such as those associated with massive energetic events in the coronae of variable stars. Dramatic shifts in intensity of these sources are typically accompanied by changes in the distribution of the wavelength of the photons emitted, known as the *spectrum* of the source. Thus, our methods aim to take advantage of changes both in the intensity of the emission and in its wavelength distribution. We illustrate our methods by applying them to high-energy observations of the flaring star known as FK Com.

Data collection in high-energy astrophysics   High-energy astrophysics is the study of electromagnetic radiation in the ultraviolet, X-ray, and γ-ray bands. The production of such high-energy photons requires temperatures of millions of degrees and signals the release of deep wells of stored energy such as those in very strong magnetic fields, extreme gravity, explosive nuclear forces, and shock waves in hot plasmas. High-energy detectors have much in common in terms of their statistical properties (and much that distinguishes them from detectors in other energy regimes). In this article we focus on statistical methods that are appropriate for high-energy astrophysics.

Astronomical data from high-energy observatories are usually obtained as a list of photons; the list records the two-dimensional direction from whence each photon arrived, the time at which it was recorded, and its wavelength (and hence its energy). Owing to the intrinsic resolutions of high-energy photon telescope/detector systems, each of these four attributes is inherently discrete and the data can be compiled into a four-way table of photon counts. Although each attribute is subject

to uncertainty (also inherent in the telescope/detector systems, e.g., the true direction of arrival is distorted by the point spread function), the quality of the resulting datasets is unprecedented in the history of astronomical observations. Nonetheless the remarkable four-way tables of photon attributes are largely an untapped resource. This is mostly due to a lack of principled statistical methods that can be used to detect and characterize astronomical sources in high-dimensional spaces. This is especially true of high-spectral-resolution grating data (e.g., data obtained using the HETGS+ACIS-S configuration of the *Chandra X-ray Observatory*), where a self-selected sample of *interesting* objects are observed for long durations (often for > 100 ksec) with Doppler resolutions of ~ 100 km s^{-1}, and with an intrinsic temporal resolution that varies from one millisecond to three seconds. Because the locations of the objects we study in this article do not change, we ignore directional information and focus on the two-way table of photon counts indexed by time and wavelength. The one-way marginal table of wavelengths is called the (observed) *spectrum* and that of the times is the (observed) *light curve* of the source.

Change points in Poisson processes   From a statistical point of view, sudden changes in an underlying physical process are modeled via change points. In a parametric model, for example, the value of the parameter is allowed to change at one or more time points during the observation. The periods of time where the parameter is fixed are called the *regimes* of the process. We might consider testing for a single change point, fitting the model with a given number of change points, or estimating the number of change points. Given that the data in high-energy astrophysics are photon counts, we focus on change points in non-homogeneous Poisson processes. Because the wavelength (and direction) of each photon are recorded, we can either model them using a marked Poisson process or count the events in each of a number of wavelength bins and model them as a multivariate Poisson process. Typical detectors have discrete temporal and spectral resolution so it is natural to use photon counts in time cross wavelength bins, i.e., to use discrete-time processes. The methods that we propose are designed specifically for two-way tables of Poisson counts of this sort.

The complexity of any statistical model naturally increases with the number of change points. Thus to avoid overfitting, methods must penalize complexity. Bayesian prior distributions and marginalization techniques are natural avenues for encouraging parsimony, for example in that they inherently embody Occam's Razor in model selection (e.g., MacKay, 2003, chapter 28). It is not surprising then that there is an extensive literature on Bayesian methods for change point models. Raftery and Akman (1986) is an early contribution for count data generated according to a homogeneous Poisson process until some unknown time when the intensity of the process changes. They propose computing the joint posterior distribution of the two Poisson rates and the single change point along with a Bayes Factor to test for existence of the change point; see Carlin *et al.* (1992) and Moreno *et al.* (2005) for derivation of hierarchal and intrinsic prior distributions, respectively, in the single change point Poisson setting. Green (1995) generalizes this approach by

using a continuous time model, allowing for multiple regimes, putting a prior distribution on the number of regimes, and fitting this number. Lai and Xing (2011), on the other hand, use a discrete time model with a Bernoulli process governing the change in regimes. This approach follows Chib (1998) who embedded a latent discrete-time discrete-state Markov process for the regime index into a multi-level Bayesian model. This formulation has been applied for inhomogeneous Poisson processes, for example, to model change points in the parameters of a log-linear model (Park, 2010), see Park *et al.* (2012) for a related approach.

Perhaps the most popular change-point method for astronomical count data is known as Bayesian Blocks (Scargle, 1998). It starts by splitting the time interval into two, assuming a homogeneous Poisson process on each, and computing the posterior odds of this model relative to a homogeneous process on the entire interval. The change point is chosen by maximizing the posterior odds and is accepted if the odds favor this model. This process continues recursively on each subinterval until no more change points are added. Scargle *et al.* (2013) improved the method by achieving global optimization over the change points while adding the capacity to handle gaps in the data, variable exposure, piecewise linear and piecewise exponential Poisson intensities, multivariate data, the analysis of variance, data on the circle, etc. Bayesian Blocks does not, however, address changes in the spectrum at the change points.

Non-Bayesian methods for change-point detection have also been proposed for non-homogeneous Poisson processes. Akman and Raftery (1986) and Worsley (1986) are early contributions. Both papers developed methods for testing for the presence of a change-point in a Poisson process. These methods follow classical frequentist arguments and also provide interval estimates for any detected change-points. Loader (1992) considered using log-linear models for non-homogeneous Poisson processes. In particular likelihood ratio tests are derived to choose a "best" model amongst different change-point and log-linear models for the observed data. In Mei *et al.* (2011) methods are proposed for detecting changes in Poisson rate after the effects of population size are taken into account. These methods are based on a generalized likelihood ratio, weighted likelihood ratio, and adaptive thresholding. More recently, Shen and Zhang (2012) applied the modified BIC (mBIC) of Zhang and Siegmund (2007) to detect change-points in non-homogeneous Poisson processes. The mBIC can be viewed as a penalized likelihood model selection criterion that is tailored to estimating the number and locations of the change points.

Change points in marked Poisson processes Despite the extensive literature on change points in Poisson processes, there is little on change points in *marked* Poisson processes. Methods have been developed to test for non-homogeneity in marked Poisson processes, for example via a scan statistic that is computed in a sliding time interval (Chan and Zhang, 2007). In high-energy astrophysics a common strategy is to compute a simple summary statistic describing the wavelength distribution in each time bin and to visually inspect a plot of the statistic as a function of time. A typical choice of statistics is the so-called *hardness ratio* which is the ratio of the photon count observed

above a given wavelength threshold to that below the threshold; see Park *et al.* (2006) for a small-sample statistical treatment of hardness ratios. Plotting hardness ratios against time is an informal exploratory technique. The primary goal of this article is to provide a change point method for marked Poisson processes that is inspired by and tailored to a specific problem in high-energy astrophysics.

Methods for detecting change points   There is a wide range of statistical approaches for detecting statistical change points. For the classical simple case of piecewise constant signals with additive noise, for example, Yao (1988) studied the use of the Schwartz criterion and showed it to be consistent. This criterion can be interpreted as the classical Bayesian information criterion (BIC) if the location of a change point is treated as a free parameter (or as a dimension of the model). In this case, however, the likelihood functions violate the regularity conditions necessary for the classical BIC, so both criteria can likely be improved. To circumvent this issue, Zhang and Siegmund (2007) provide an ingenious method to approximate the Bayes factor and developed the above-mentioned mBIC for change point detection which has been shown to be consistent in various settings. More recently, Ninomiya (2014) derived a specialized Akaike information criterion (AIC) for change point estimation. Results from Aue and Lee (2011, Section 3.3), however, can be used to show that this AIC method is not consistent.

In this article we focus on the minimum description length (MDL) principle (Rissanen, 1989, 2007), which can also be viewed as a penalized likelihood approach to model selection. A brief introduction of MDL is given in Section 3.3. One of its earliest applications to change point estimation is the image segmentation work of Leclerc (1989), where the corresponding MDL criterion can be easily reduced to the classical one-dimensional change point problem setting; see also Lee (1998) for an extension to correlated noise. Other successful MDL applications to change point problems include discontinuity detection in nonparametric regression (Lee, 2002a), structural break detection in nonstationary time series models (Davis and Yau, 2013; Yau *et al.*, 2015), and change point detection in quantile modeling (Aue *et al.*, 2014). Under mild regularity assumptions, these MDL solutions can be shown to be consistent.

For many change-point problems simpler than the one considered in this paper, the major penalty terms of mBIC and MDL are remarkably similar; e.g., compare Equation (8) of Zhang and Siegmund (2007) and Equation (6.2) of Lee (1997). This provides reassuring evidence that both mBIC and MDL are reliable methods for change point detection.

Strategy and outline   We pursue a non-parametric penalized-likelihood strategy, using MDL to optimize the tuning parameters of the penalty function and to fit the number of change points. Between the change points, we assume a homogeneous Poisson process where the spectrum includes a broad smooth term that is modeled using a radial basis expansion. In particular the bases we use are cubic polynomials, which provide a good balance between model flexibility and ease of

implementation. We also include a number of single-bin local deviations from this extended smooth model. Physically, these deviations correspond to spectral emission lines that are the result of the production of photons of a particular wavelength as electrons fall to the lower energy shells of a specific ion. Both the radical basis expansion and the local deviations are over parameterized in that we expect most of their coefficients and intensities (respectively) to be zero. Thus, we employ an ℓ_1 penalty on both. An R package "Automark" implementing our proposed methodology is available to interested readers at `https://github.com/astrostat/Automark`.

Some simple spectral models such as power laws and exponential absorption can be formulated as log-linear models (van Dyk *et al.*, 2001) and thus in principle could be embedded into a temporal model with change points for their parameters, similar to the proposal of Park (2010). We take a different approach by using flexible non-parametric spectral models. This allows us to account for both blurring of the recorded energies and background contamination of the photon counts, and to apply a single model irrespective of the particular shape of the source spectra. (See van Dyk and Kang (2004) for a review of the source and data-distortion models used in high-energy spectral analysis.) This strategy is contrary to the typical approach in high-energy spectral analysis where physics-based parametric models are preferred because their parameters have direct scientific meaning. Our primary goal here, however, is to identify the change points. The spectra corresponding to each regime within the process can be fit using physics-based models in a secondary analysis.

This article is organized into seven sections. Section 2 describes the detectors used for data collection, the relevant astrophysical models, and basic notation for our statistical approach. We formalize our model for the homogeneous Poisson processes within each regime of the overall process in Section 3 and describe how we embed this into the change-point model in Section 4. Numerical results including a set of simulation studies and an application to the yellow giant FK Com appear in Sections 5 and 6, respectively. Discussion appears in Section 7 and details of the derivation of our MDL criterion in an Appendix.

2   Astrophysical Background and Statistical Notation

Suppose that we observe photon counts in an $N \times J$ rectangular array of equally-sized wavelength-cross-time bins, where y_{ij} is the photon count in the bin with wavelength range $[w_i - \delta_w/2, w_i + \delta_w/2)$ and temporal range $[t_j - \delta_t/2, t_j + \delta_t/2)$, for $i = 1, \ldots, N$ and $j = 1, \ldots, J$. Although this array may be compiled using the natural resolution of the detector, with high-resolution data it may be computationally advantageous to combine the natural bins in order to obtain a lower resolution array of photon counts. This strategy is illustrated and discussed in Section 6. In any case, the expected count in each bin is ideally proportional to the brightness of the astronomical source integrated over its time and wavelength ranges. Because of detector effects, however, a photon with wavelength w has a distribution of possible recorded energies, w'. This distribution is called the *detector response function* and denoted $R(w, w')$. We apply our methods to data collected using

a *grating* that, like a prism, spatially separates photons as a function of their wavelength. With grating data, wavelength is measured very accurately and we can ignore measurement uncertainty and treat $R(w, w')$ as a Dirac delta function. Another detector effect arises because the sensitivity of the detector varies with photon wavelength, an effect that is quantified via the so-called *effective area* of the detector and denoted by $A(w)$. The effective area quantifies characteristics of the telescope mirrors that do not relate to the nature of the source and is typically treated as known. (See, however, Lee *et al.* (2011) and Xu *et al.* (2014) for development of Bayesian statistical methods that account for uncertainty in $A(w)$.)

Using a Poisson model for the photon counts and letting $\lambda(t_j, w_i)$ be the expected count per unit time and per unit wavelength averaged over the bin centered at (t_j, w_i),

$$y_{ij} \stackrel{\text{ind}}{\sim} \text{Poisson}\{\delta_t \cdot \delta_w \cdot \lambda(t_j, w_i) \cdot A(w_i)\}. \tag{1}$$

In practice, we may combine data from K detectors, each with its own effective area. Letting $A_1(w), \ldots, A_K(w)$ denote the K effective areas, the total counts across the K detectors can be modeled

$$Y_{ij} = \sum_{k=1}^{K} y_{ijk},$$

where

$$y_{ijk} \stackrel{\text{ind}}{\sim} \text{Poisson}\{\delta_t \cdot \delta_w \cdot \lambda(t_j, w_i) \cdot A_k(w_i)\}$$

are the counts recorded with detector k. Because the y_{ijk} are independent across k, we have

$$Y_{ij} \sim \text{Poisson}\left\{\delta_t \cdot \delta_w \cdot \lambda(t_j, w_i) \sum_{k=1}^{K} A_k(w_i)\right\}.$$

The goal of this article is to infer the properties of $\lambda(t_j, w_i)$ using the observed photon counts, $\{Y_{ij}\}$. We are particularly interested in detecting statistically significant changes in $\{Y_{ij}\}$ over time. In other words, we would like to determine if there are any change point, π, such that $\lambda(t_j, w_i)\big|_{\{t_j \leq \pi\}}$ differs from $\lambda(t_j, w_i)\big|_{\{t_j > \pi\}}$. If there are, we aim to estimate the number of change points and their values.

For ease of presentation, we develop our estimation procedure in two stages. First we develop a flexible model for a time-homogeneous source spectrum. That is we assume that there are no change points and that $\lambda(t_j, w_i)$ is constant in t_j. In the second stage we introduce change points and allow $\lambda(t_j, w_i)$ to change over time.

7

3 A Model for the Homogeneous Poisson Processes

3.1 A nonparametric spectral model

Because there are no change points in our time-homogeneous Poisson model, we drop t_j from the argument of $\lambda(t_j, w_i)$ and write $\lambda(t_j, w_i) = \lambda(w_i)$ in this section. Also, we denote the expectation of Y_{ij} by $\mu(w_i)$. That is, $\mu(w_i) = s(w_i)\lambda(w_i)$ with $s(w_i) = \delta_t \cdot \delta_w \sum_{k=1}^{K} A_k(w_i)$ being a completely specified function. We represent the expected counts due to the underlying physical process by λ and expected detector counts by μ; μ is adjusted for varying bin sizes and effective areas.

The parameter of scientific interest is $\lambda(w_i)$. For the astronomical sources that we study $\lambda(w_i)$ is mostly smooth with a few emission lines overlaid and we adopt the model

$$\lambda(w_i) = f(w_i) + \sum_{m=1}^{M} \alpha_m g_m(w_i, \tau_m), \tag{2}$$

where $f(w)$ is a nonnegative smooth function, M is the number of emission lines, $\alpha_m > 0$ is the magnitude of emission line m, and $g_m(w, \tau_m)$ models the shape of the emission line centered at τ_m. Common choices for $g_m(w, \tau_m)$ include Gaussian density functions with small variances, Lorentzian density functions, and delta functions (e.g., van Dyk *et al.*, 2006, and the references therein). Here we assume that each of the emission lines locates completely within a single bin, although our framework can be easily modified for cases when some emission lines span multiple bins. For astrophysical reasons, we also assume that each $\alpha_m > 0$, that is each emission line constitutes a positive deviation from $f(w)$, although in principle the α_m may also be negative, so long as $\lambda(w_i) > 0$.

If an individual spectral line is located in a single bin, the line profile, $g_m(w, \tau_m)$, is an indicator function for the interval of width δ_w centered at $\tau_m \in \{w_1, \ldots, w_N\}$, that is, an indicator function for one of the wavelength bins. The natural resolution of the bin counts does not allow modeling of $g_m(w, \tau_m)$ at any greater level of detail than this. Under this model for the emission lines, $\lambda(w_i)$ in (2) becomes

$$\lambda(w_i) = f(w_i) + \sum_{m=1}^{M} \alpha_m I\left(\tau_m - \frac{\delta_w}{2} \leq w_i < \tau_m + \frac{\delta_w}{2}\right),$$

where $I(\mathcal{A})$ denotes the indicator function for $\mathcal{A}$ and $\tau_m \in \{w_1, \ldots, w_N\}$. Because of the nonnegativity constraint on $\lambda(w_i)$, we introduce a log-link function,

$$\log \lambda(w_i) = \log f(w_i) + \sum_{m=1}^{M} \alpha'_m I\left(\tau_m - \frac{\delta_w}{2} \leq w_i < \tau_m + \frac{\delta_w}{2}\right) \tag{3}$$

where α'_m is the transformed α_m that ensures the equality in (3) holds.

To provide flexible modeling, we use a radial basis expansion (see, e.g., Buhmann, 2003; Ruppert

et al., 2003) to nonparametrically model $\log f(w)$, the smooth component of $\log \lambda(w)$. Specifically, we use the P basis functions, $(\xi_1(w), \ldots, \xi_P(w))$, which are polynomials of power 3 with equally spaced knots, $\kappa_1, \ldots, \kappa_{P-4}$.[1] Specifically, the basis functions are

$$\xi_1(w) = 1, \quad \xi_2(w) = w, \quad \xi_3(w) = w^2, \quad \xi_4(w) = w^3, \quad \text{and}$$
$$\xi_{p+4}(w) = |w - \kappa_p|^3, \qquad p = 1, \ldots, P - 4.$$

Substituting these into (3) yields

$$\log \lambda(w_i) = \sum_{p=1}^{P} \beta_p \xi_p(w_i) + \sum_{m=1}^{M} \alpha'_m I\left(\tau_m - \frac{\delta_w}{2} \le w_i < \tau_m + \frac{\delta_w}{2}\right), \tag{4}$$

where β_p is the coefficient of basis function p.

Model (4) is an example of a generalized semi-parametric model. Under this formulation, estimating $\lambda(w_i)$ is equivalent to estimating P, $(\beta_1, \ldots, \beta_P)$, M, and $(\alpha'_1, \tau_1, \ldots, \alpha'_M, \tau_M)$.

3.2 A lasso model for the emission lines

A difficulty in fitting model (4) is the estimation of $(\tau_1, \ldots, \tau_M)$, the locations of the emission lines. With M unknown, there is a total of 2^N possibilities for $(\tau_1, \ldots, \tau_M)$. However, as emission lines are relatively rare (i.e., sparse), we can take advantage of the celebrated ℓ_1 penalty techniques (e.g., Tibshirani, 1996) to provide a fast searching algorithm for $(\tau_1, \ldots, \tau_M)$.

We begin by rewriting (4) as

$$\log \lambda(w_i) = \sum_{p=1}^{P} \beta_p \xi_p(w_i) + \sum_{k=1}^{N} \eta_k I_k(w_i) = \sum_{p=1}^{P} \beta_p \xi_p(w_i) + \eta_i, \tag{5}$$

with $I_k(w) = I(w_k - \delta_w/2 \le w < w_k + \delta_w/2)$. In this formulation $\eta_i = 0$ implies there is no emission line in the wavelength bin centered at w_i. There is a one-to-one correspondence between models (4) and (5) via

$$M = \sum_{i=1}^{N} I(\eta_i \ne 0) \quad \text{and} \quad \left\{(\tau_m, \alpha'_m)\right\}_{m=1}^{M} = \left\{(w_i, \eta_i) : \eta_i \ne 0\right\}.$$

A major advantage of re-expressing model (4) as (5) is that, many of the $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_N)^\mathsf{T}$ are zero in (5). Therefore by imposing an ℓ_1 penalty, we can achieve fast simultaneous selection and estimation of the nonzero elements among $\boldsymbol{\eta}$.

We employ a common strategy for the smooth component, $\log f(w_i) = \sum_{p=1}^{P} \beta_p \xi_p(w_i)$, in (5).

---

[1] The gap between the first knot the left endpoint of the spectral range and the gap between the last knot and the right endpoint of the range are both slightly larger than the gaps between consecutive knots.

Namely, we specify a large value of P *a priori* and estimate $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_P)^\intercal$ under a penalty term that aims to avoid over-fitting. In this way, the coefficients of the the redundant functions among $\xi_1(w), \ldots, \xi_P(w)$ are shrunk toward zero. This enables us to assume that P is known (and large).

Penalized maximum likelihood can be used to simultaneously estimate all of the unknown parameters in (5), including those describing the smooth component and those describing the emission lines.

For a Poisson random variable with expectation $a > 0$, we write the corresponding probability mass function as

$$q(x; a) = -a + x \log a - \log(x!) \tag{6}$$

for any nonnegative integer x. The log-likelihood of a bin count Y_{ij} is

$$l_{\text{one}}(Y_{ij}; \boldsymbol{\theta}) = q\{Y_{ij}; \mu(w_i)\} = q\{Y_{ij}; s(w_i)\lambda(w_i)\},$$

where $\lambda(w_i)$ is parameterized in terms of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\eta})$ as in (5). (This likelihood will be used within one homogenous time segement in Section 4; hence the subscript "one".) Adding a penalty, we define the estimate of $\boldsymbol{\theta}$ as the maximizer of

$$\sum_{i=1}^{N} \sum_{j=1}^{J} l_{\text{one}}(Y_{ij}; \boldsymbol{\theta}) - \gamma\{\rho\|\mathbf{D}\boldsymbol{\beta}\|_1 + (1 - \rho)\|\boldsymbol{\eta}\|_1\}, \tag{7}$$

where $\mathbf{D} = \text{diag}\{(0, 0, 0, 0, 1, \ldots, 1)\}$ is a diagonal matrix, and $\gamma > 0$ and $0 \leq \rho \leq 1$ are tuning parameters that determine the penalty on $\boldsymbol{\theta}$. The first four diagonal elements of $\mathbf{D}$ are set to zero because there is no penalty for the inclusion of $(\beta_1, \ldots, \beta_4)$ in the model. If γ and ρ are specified, maximization of (7) is equivalent to lasso fitting of a generalized linear model. There are fast algorithms for this optimization problem (e.g., coordinate descent, Friedman *et al.*, 2010) and software is widely available; e.g., the R package `glmnet`.

3.3   Tuning parameter selection using MDL

The success of (7) depends heavily on our ability to choose good values of γ and ρ, which determine the number of emission lines and the number of basis functions used in the smooth component of the fitted model. To select γ and ρ, we use the minimum description length (MDL) principle (Rissanen, 1989, 2007). In short, MDL defines the best model as the one that achieves the highest compression rate of the data, $\mathcal{D}$. In other words, the best model allows us to store $\mathcal{D}$ with the shortest codelength. A good statistical model and a good compression method share a common feature: they should be able to capture regularities and patterns hidden in the data. Therefore it is reasonable to expect that a model chosen by MDL to be a good statistical model; some successful examples for change point problems were provided in Section 1, while for other applications can be found in the later chapters of Grünwald *et al.* (2005). There are various versions of MDL. We use

the so-called two-part form of MDL to derive a model selection criterion for choosing γ and ρ.

Let $\mathsf{CL}(\mathcal{Z})$ be the codelength of $\mathcal{Z}$: one can treat $\mathsf{CL}(\mathcal{Z})$ as the amount of memory needed to store $\mathcal{Z}$. Similarly, let $\mathsf{CL}(\mathcal{Z}|\mathcal{X})$ be the codelength of $\mathcal{Z}$ conditional on $\mathcal{X}$, that is, the codelength of $\mathcal{Z}$ with knowledge of $\mathcal{X}$. Generally speaking, the two-part form of MDL decomposes the codelength of the data $\mathcal{D}$ into two parts

$$\mathsf{CL}(\mathcal{D}) = \mathsf{CL}(\widehat{\mathcal{M}}) + \mathsf{CL}(\mathcal{D}|\widehat{\mathcal{M}}), \tag{8}$$

where the best fitting model $\widehat{\mathcal{M}}$ is defined to be the minimizer of $\mathsf{CL}(\mathcal{D})$. In Appendix A it is shown that when fitting model (7), for large N the two-part MDL (8) is asymptotically equal to

$$\mathsf{mdl}_{\mathrm{null}}(\gamma, \rho) = -\sum_{i=1}^{N}\sum_{j=1}^{J} l_{\mathrm{one}}(Y_{ij}; \hat{\boldsymbol{\theta}}) + \frac{1}{2}\|\hat{\boldsymbol{\theta}}\|_0 \log(NJ) + \log\binom{N}{\|\hat{\boldsymbol{\eta}}\|_0}, \tag{9}$$

where $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\eta}})$ is the maximizer of (7) given γ and ρ, and the ℓ_0 norm, $\|\boldsymbol{z}\|_0$, denotes the number of nonzero elements in the vector $\boldsymbol{z}$. We choose the values of γ and ρ that jointly maximize (9) by evaluating $\mathsf{mdl}_{\mathrm{null}}(\gamma, \rho)$ on a fine grid, using $\texttt{glmnet}$ to compute $\hat{\boldsymbol{\theta}}(\gamma, \rho)$ on a grid of values of γ for each value of ρ. In many MDL applications, a term similar to the last one in (9) is negligible and hence omitted. Here, however, the number of unknown parameters $(N + P)$ is not ignorable when compared to the number of observations and hence the last term in (9) is required.

In summary, for the time-homogeneous model with no change points, the parameter estimate $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\eta}})$ is obtained by maximizing (7) with γ and ρ chosen as the joint minimizer of (9). Finally, the estimate for $\lambda(w_i)$ is calculated as $\hat{\lambda}(w_i) = \exp\{\sum_{p=1}^{P} \hat{\beta}_p \xi_p(w_i) + \hat{\eta}_i\}$.

4 Modeling with Change Points

4.1 Adding change points to the spectral model

In this section, we allow the spectrum, $\lambda(t_j, w_i)$, to change over time. Hence we reinstate the notation that emphasized the dependence on t_j. We model $\lambda(t_j, w_i)$ as piecewise constant as a function of t_j. This involves partitioning the entire time interval $[0, T)$ into several subintervals and independently modeling $\lambda(t_j, w_i)$ as constant in t_j in each subinterval using a model of the form given in (5). Again let Y_{ij} be the photon count in the bin centered at (w_i, t_j) and summed over the K detectors. We can formalize the change-point model for $\lambda(t_j, w_i)$ as

$$\lambda(t_j, w_i) = \sum_{b=1}^{B} I_b(t_j) \lambda_b(w_i), \tag{10}$$

where $B \geq 1$ is the number of time (sub)intervals, $I_b(t_j) = I(\pi_{b-1} \leq t_j < \pi_b)$ identifies the time bins in interval b, $t_1 - \delta_t/2 = \pi_0 < \cdots < \pi_B = t_J + \delta_t/2$ are the change points, and $\lambda_b(w_i)$ is the

spectrum in time interval b. For ease of discussion, we refer to $\pi_1, \ldots, \pi_{B-1}$ as the *change points*, excluding the endpoints π_0 and π_B. The within time-interval spectra are modeled as in (5), i.e.,

$$\log \lambda_b(w_i) = \sum_{p=1}^{P} \beta_{bp} \xi_p(w_i) + \eta_{bi},$$

where $\boldsymbol{\beta}_b = (\beta_{b1}, \ldots, \beta_{bP})^{\mathsf{T}}$ and $\boldsymbol{\eta}_b = (\eta_{b1}, \ldots, \eta_{bN})^{\mathsf{T}}$ are the parameters for the time-interval-specific spectra.

There is no gain in allowing the change points, $(\pi_1, \ldots, \pi_{B-1})$, to take values other than bin breaks if there is no prior knowledge about the change points. Further, to maintain sufficient data within each time interval for acceptable estimation, $|\pi_i - \pi_j|$ cannot be too small for any $0 \le i, j \le B$. We require that $|\pi_i - \pi_j| \ge 5\delta_t$ in our numerical illustrations. This ensures that each time interval is at least five bins wide.

4.2 Model selection using MDL

In order to fit model (10), we must estimate the number of time intervals, B, the change points $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_{B-1})^{\mathsf{T}}$, and the parameters for each interval, $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_B\}$, where $\boldsymbol{\theta}_b = (\boldsymbol{\beta}_b, \boldsymbol{\eta}_b)$. Once B and $\boldsymbol{\pi}$ are specified, however, each $\lambda_b(w_i)$ along with its tuning parameters, $\gamma = \gamma_b$ and $\rho = \rho_b$, can be estimated individually using the method of Section 3. Thus, we aim to first estimate B, $\boldsymbol{\pi}$ and then compute $\hat{\gamma}_b(B, \boldsymbol{\pi})$, $\hat{\rho}_b(B, \boldsymbol{\pi})$, and $\hat{\boldsymbol{\theta}}_b(\hat{\gamma}_b, \hat{\rho}_b)$ as in Section 3 for $b = 1, \ldots, B$. This notation signifies the dependence of $\hat{\gamma}_b$ and $\hat{\rho}_b$ on $(B, \boldsymbol{\pi})$, and the dependence of $\hat{\boldsymbol{\theta}}_b$ on $(\hat{\gamma}_b, \hat{\rho}_b)$. Similarly, we also write $\boldsymbol{\Theta}$ and $\lambda(t_j, w_i)$ as $\boldsymbol{\Theta}(\boldsymbol{\gamma}, \boldsymbol{\rho})$ and $\lambda \{t_j, w_i; B, \boldsymbol{\pi}, \boldsymbol{\Theta}(\boldsymbol{\gamma}, \boldsymbol{\rho})\}$, respectively.

The profile penalized (pp) loglikelihood function for B, $\boldsymbol{\pi}$ under model (10) depends on the tuning parameters $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_B)$ and $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_B)$ and can be written

$$\log \mathcal{L}_{\mathsf{pp}}(B, \boldsymbol{\pi}; \boldsymbol{\gamma}, \boldsymbol{\rho}) = \sum_{i=1}^{N} \sum_{j=1}^{J} q \left(Y_{ij}; s(w_i) \lambda \left\{ t_j, w_i; B, \boldsymbol{\pi}, \widehat{\boldsymbol{\Theta}}(\boldsymbol{\gamma}, \boldsymbol{\rho}) \right\} \right), \tag{11}$$

with q the Poisson loglikelihood given in (6). Because $\boldsymbol{\gamma}$ and $\boldsymbol{\rho}$ can be computed as functions of B and $\boldsymbol{\pi}$, we can substitute $\hat{\boldsymbol{\gamma}}(B, \boldsymbol{\pi})$ and $\hat{\boldsymbol{\rho}}(B, \boldsymbol{\pi})$ into $\log \mathcal{L}_{\mathsf{pp}}(B, \boldsymbol{\pi}; \boldsymbol{\gamma}, \boldsymbol{\rho})$. This is akin to profiling over $\boldsymbol{\gamma}$ and $\boldsymbol{\rho}$, except that $\hat{\boldsymbol{\gamma}}(B, \boldsymbol{\pi})$ and $\hat{\boldsymbol{\rho}}(B, \boldsymbol{\pi})$ are computed according to the MDL criterion rather than by maximizing $\log \mathcal{L}_{\mathsf{pp}}(B, \boldsymbol{\pi}; \boldsymbol{\gamma}, \boldsymbol{\rho})$. The result is the pseudo profile penalized (ppp) loglikelihood given by

$$\begin{aligned} \log \mathcal{L}_{\mathsf{ppp}}(B, \boldsymbol{\pi}) &= \log \mathcal{L}_{\mathsf{pp}}(B, \boldsymbol{\pi}; \hat{\boldsymbol{\gamma}}(B, \boldsymbol{\pi}), \hat{\boldsymbol{\rho}}(B, \boldsymbol{\pi})) \\ &= \sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{b=1}^{B} I_b(t_j) l_{\mathrm{one}} \left\{ Y_{ij}; \hat{\boldsymbol{\theta}}_b(B, \boldsymbol{\pi}) \right\}, \end{aligned} \tag{12}$$

12

where $\hat{\boldsymbol{\theta}}_b(B, \boldsymbol{\pi}) = \hat{\boldsymbol{\theta}}_b\big(\hat{\boldsymbol{\gamma}}(B, \boldsymbol{\pi}), \hat{\boldsymbol{\rho}}(B, \boldsymbol{\pi})\big)$. Because different values of B lead to different numbers of parameters in the model, we cannot estimate the parameters by maximizing (12). Instead we view this as a model selection problem and again use the MDL principle. In this way we can choose the "best" combination of B and $\boldsymbol{\pi}$, and then compute $\hat{\gamma}_b(B, \boldsymbol{\pi})$, $\hat{\rho}_b(B, \boldsymbol{\pi})$, and $\hat{\boldsymbol{\theta}}_b(\hat{\gamma}_b, \hat{\rho}_b)$ for each time interval. We derive the MDL criterion in Appendix B and show that the best model can be found by minimizing

$$\mathsf{mdl}_{\mathrm{full}}(B, \boldsymbol{\pi}) = -\log \mathcal{L}_{\mathsf{ppp}}(B, \boldsymbol{\pi}) + \log B + \sum_{b=1}^{B} \log c_b(\boldsymbol{\pi})$$

$$+ \sum_{b=1}^{B} \left[\frac{1}{2} \|\hat{\boldsymbol{\theta}}_b(B, \boldsymbol{\pi})\|_0 \log\{N c_b(\boldsymbol{\pi})\} + \log \binom{N}{\|\hat{\boldsymbol{\eta}}_b(B, \boldsymbol{\pi})\|_0} \right], \tag{13}$$

where $c_b(\boldsymbol{\pi})$ is the cardinality of $\{i : \pi_{b-1} \le t_i < \pi_b\}$ for $b = 1, \ldots, B$, i.e., $c_b(\boldsymbol{\pi})$ is the number of time bins in time interval b.

4.3 Practical minimization

We now consider minimization of the MDL criterion given in (13). Once the number of time intervals, B, and the locations of change points, $\boldsymbol{\pi}$, are specified, unique estimates of $\boldsymbol{\gamma}$ and $\boldsymbol{\rho}$ can be obtained using the method described in Section 3. Since the time intervals are independent the required computation is highly parallelizable. Nonetheless, minimization of (13) is not trivial. It involves an expensive combinatoric optimization: for each B, there are $\binom{J-1}{B-1}$ possibilities for $\boldsymbol{\pi}$. (A small number of these possibilities is not considered by our optimiziation procedure because we impose a minimum width on each time interval.) To simplify computation, we propose a fast stepwise forward algorithm to approximately minimize (13), rather than attempting to globally minimize it. Briefly, at each step we add a single change point to the current collection by selecting the new change point that decreases $\mathsf{mdl}_{\mathrm{full}}(B, \boldsymbol{\pi})$ the most.

More precisely, the algorithm first calculates (13) for the model with no change points (i.e., only one time interval), that is, it computes $\mathsf{mdl}_1 = \mathsf{mdl}_{\mathrm{full}}(1, \boldsymbol{\pi})$. (This does not depend on $\boldsymbol{\pi}$ because there are no change points.) In the second step, the algorithm considers all of the possible models with a single change point and selects the one that minimizes $\mathsf{mdl}_{\mathrm{full}}(2, \boldsymbol{\pi})$. Calling the obtained minimum mdl_2, the change point is added to the model if $\mathsf{mdl}_2 < \mathsf{mdl}_1$; otherwise the algorithm terminates. The algorithm continues in this way, attempting to add an additional change point to the model at each step by minimizing (13). It stops when there is no possible change point remaining that decreases $\mathsf{mdl}_{\mathrm{full}}(B, \boldsymbol{\pi})$ and this current model is taken as the (approximate) minimizer of (13). Although this greedy algorithm may result in local minimization, it is fast and provides a good compromise between computational efficiency and accuracy (Lee, 2002b). The exact steps of the algorithm appear in Algorithm 1.

Algorithm 1 A stepwise search algorithm for minimizing (13)

Input: Wavelength-cross-time bin counts $\{(Y_{ij})\}_{i=1,\ldots,N, j=1,\ldots,J}$
Description: A fast algorithm for approximately minimizing $\mathsf{mdl}_{\mathrm{full}}(B, \boldsymbol{\pi}, \boldsymbol{\gamma}, \boldsymbol{\rho})$

  1: $B \leftarrow 1$ and $\boldsymbol{\pi} \leftarrow ()$ (null vector)
  2: $z \leftarrow \mathsf{mdl}_{\mathrm{full}}(1, \boldsymbol{\pi})$.
  3: $z' \leftarrow z + 1$
  4: **while** $z < z'$ **do**
  5:     $B \leftarrow B + 1$
  6:     $z' \leftarrow z$
  7:     $\boldsymbol{\pi}' \leftarrow \boldsymbol{\pi}$
  8:     Use a grid search to find the change point that when added to the current collection, $\boldsymbol{\pi}'$, gives the largest reduction in $\mathsf{mdl}_{\mathrm{full}}$. Denote this new set of change points as $\boldsymbol{\pi}$. If there is no possible change point to add, break the while-loop.
  9:     $z \leftarrow \mathsf{mdl}_{\mathrm{full}}(B, \boldsymbol{\pi})$
10: **end while**
11: Return $\{B - 1, \boldsymbol{\pi}'\}$

4.4 A Monte Carlo Procedure for Testing for the Existence of Change Points

In practice it is useful to quantify evidence for existence of the change points. For a test statistic, we use the change in the MDL function (13) due to introduction of change points, $m^{\star} = \mathsf{mdl}_{\mathrm{full}}(\hat{B}, \hat{\boldsymbol{\pi}}) - \mathsf{mdl}_{\mathrm{full}}(1, \boldsymbol{\pi}_{\mathrm{null}})$, where $\hat{B}$ and $\hat{\boldsymbol{\pi}}$ are computed by minimizing (13), and $\boldsymbol{\pi}_{\mathrm{null}}$ is the null vector containing no change point. Here $\mathsf{mdl}_{\mathrm{full}}(1, \boldsymbol{\pi}_{\mathrm{null}})$ is the value of the MDL function (13) with no change points. We approximate the null distribution of $m^{\star}$ via Monte Carlo using permutation. Specifically, we generate N_{sim} uniformly random permutations of $\{1, 2, \ldots, J\}$ independently and the corresponding replicate null datasets, $(\mathcal{D}_1, \ldots, \mathcal{D}_{N_{\mathrm{sim}}})$, by shuffling the time indices of the original dataset according to these permutations (see Step 3 of Algorithm 2). Each replicate dataset is then fit with the full model (including change points) according to the MDL criterion specified in (13) and the resulting test statistic, $m_i^{\star}$ is computed (for $i = 1, \ldots, N_{\mathrm{sim}}$). Note that the MDL criterion (13) with no change points, $\mathsf{mdl}_{\mathrm{full}}(1, \boldsymbol{\pi}_{\mathrm{null}})$, is invariant to the permutation and thus need not to be re-computed for the permuted datasets $(\mathcal{D}_1, \ldots, \mathcal{D}_{N_{\mathrm{sim}}})$. The empirical distribution of $(m_1^{\star}, \ldots, m_{N_{\mathrm{sim}}}^{\star})$ is a Monte Carlo approximation to the null distribution of $m^{\star}$ and a p-value can be computed by comparing $m^{\star}$ to $(m_1^{\star}, \ldots, m_{N_{\mathrm{sim}}}^{\star})$. Details appear in Algorithm 2.

5 Numerical Experiments

We conducted a series of simulation experiments to evaluate the empirical properties of the proposed methodology. Binned photon counts were simulated under eight different spectral-temporal test functions, $\lambda(t_j, w_i)$. In order to make the experiments as realistic as possible, seven of the test

---

Algorithm 2 Monte Carlo significance test

Input: Working data $\{(w_i, t_j, Y_{ij})\}_{i=1,\ldots,N,j=1,\ldots,J}$, number of simulation N_{sim}, significance level α
Description: Monte Carlo test for change point

1: Fit the time homogenous Poisson moded (with no change point, Section 3) and full Poisson model (including change points, Section 4.3) to $\{(w_i, t_j, Y_{ij})\}_{i=1,\ldots,N,j=1,\ldots,J}$ and denote the test statistic $m^\star = \mathsf{mdl}_{\text{full}}(\hat{B}, \hat{\boldsymbol{\pi}}) - \mathsf{mdl}_{\text{full}}(1, \boldsymbol{\pi}_{\text{null}})$. ($\boldsymbol{\pi}_{\text{null}} = ()$, null vector)
2: **for** $k = 1$ to N_{sim} **do**
3:     Generate the data set $\mathcal{D}_k = \{(w_i, t_j, Y_{ip_j})\}$ where $\{p_1, \ldots, p_J\}$ is a uniformly random permutation of $\{1, 2, \ldots, J\}$.
4:     Fit the full Poisson model (including change points) using the procedure in Section 4.3 to $\mathcal{D}_k$.
5:     Denote the resulting test statistic by $m_k^\star$.
6: Estimate the p-value as
$$\hat{p} = \frac{\#\{m_k^\star \geq m^\star\} + 1}{N_{\text{sim}} + 1}$$
7: Conclude that there is at least one change point if $\hat{p} < \alpha$ for some significance level, α.

---

functions were constructed by fitting a spectral-temporal model, using the method described in Section 4, to the seven observations of FK Com described in Section 6. Because these test functions have at most two change points and because we would like to investigate the statistical properties of our method with more change points, an eighth test function was created by concatenating two of the original test functions. (I particular, those created from observation identification (ObsID) numbers 13251 and 12298, see Table 1.) The analyses carried out here, including the binning of the data, is identical to that described in Section 6. Further details are summarized in Table 1. Together these eight test functions cover a wide range of possible scenarios, ranging from models with no change point to models with four change points. For each of the eight $\lambda(t_j, w_i)$, 200 data sets were simulated according to (1), using the real data values for δ_t, δ_w, and $A(w_i) = \sum_{k=1}^{K} A_k(w_i)$. The proposed methodology was applied to all $8 \times 200 = 1400$ simulated data sets to estimate $\lambda(t_j, w_i)$ for each.

The simulation results are summarized in Table 2. The first column lists the ObsID number for the observations of FK Com (in chronological rather than ascending order) used to generate the $\lambda(t_j, w_i)$, see Section 6. The second column shows the (true) numbers of homogeneous time intervals, B, in each of the $\lambda(t_j, w_i)$; recall the number of change points is $B - 1$. The third column reports the percentage of the simulated datasets for which the fitted $\hat{B}$ exactly equals the true value. The proposed methodology achieved a correct recovery rate of at least 94% in 6 out of the 8 simulations settings.

Our method also accurately estimates the change points, $\boldsymbol{\pi}$, as shown in Column 4 of Table 2. For those simulations where B is correctly estimated, Column 4 gives the root mean squared error

Table 1: Details of simulation settings. The columns report (1) the chronologically ordered (except the last row) ObsID numbers of the datasets used to generate each test function, (2) the number of wavelength bins, (3) the number of time bins, (4) the number of basis functions P, as in Equation (4), used to fit the corresponding FK Com observation, (5) the number of time intervals, i.e., the number of change points plus one, estimated from the corresponding FK Com observation, (6) vector of the numbers of emission lines with element j corresponding to time interval j, estimated from the corresponding FK Com observation.

ObsID	N	J	P	B	M
12297	142	21	34	2	(0, 1)
12356	142	32	34	1	(1)
13251	142	49	34	3	(0, 0, 0)
12298	142	20	34	2	(2, 1)
13259	142	23	34	2	(0, 0)
12357	142	18	34	1	(1)
12299	142	20	34	3	(0, 0, 0)
13251, 12298	142	69	34	5	(0, 0, 0, 2, 1)

Table 2: Simulation results. The columns report (1) the chronologically ordered (except the last row) ObsID numbers of the datasets used to generate each test function, (2) the true number of time intervals, i.e., the true number of change points plus one, (3) the percentage of the fitted $\hat{B}$ that exactly equal the true B, and (4) the root mean square error of $\hat{\pi}$ among those simulations with $\hat{B} = B$.

ObsID	B	% correct $\hat{B}$	RMSE($\hat{\pi}$)
12297	2	94	0.70
12356	1	100	0
13251	3	98	0.66
12298	2	98	0.29
13259	2	80	0.69
12357	1	100	0
12299	3	100	0
13251, 12298	5	85	0.66

(RMSE) of $\hat{\pi}$, in terms of the time bin width ($\delta_t = 2000$ seconds in all cases). The RMSE is less than one bin width in all eight cases. Figure 1 displays histograms of the estimated change point location among those simulations for which $\hat{B} = B$ under the five test functions with less than a 100% recovery rate for π. In all five cases, the estimated change point locations are narrowly centered around the true change points. Finally, for those simulations for which $\hat{B} = B$ and $\hat{\pi} = \pi$, the uncertainty of $\hat{f}$ is summarized in Figure 2. Here we only present results for one test function (ObsID 12299); plots for the other test functions appear in Appendix C. For each of the $B = 3$ homogeneous time intervals associated with ObsID 12299, we overlay the fitted $\hat{f}$ of the simulations for which $\hat{B} = B$ and $\hat{\pi} = \pi$. The results are plotted in panels (a)-(c) of Figure 2. The relatively high uncertainty at high wavelengths is partly due to the small effective area at these wavelengths; see panel (d) of Figure 2.
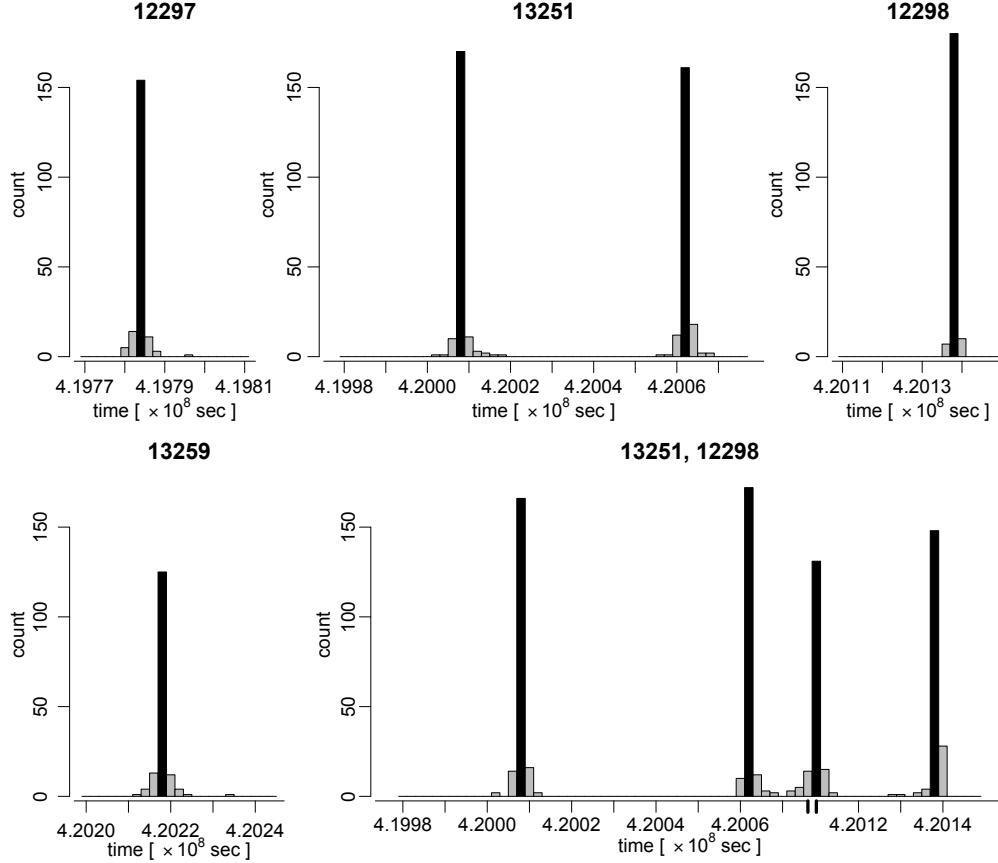
Figure 1: Histograms of detected change point locations in the simulation study. The title of each panel refers to the ObsIDs given in Tables 1 and 2. The widths of the histogram bins is the same as the time binning of the data; bins containing the true change-points locations are represented in black. The horizontal axis is elapsed spacecraft time and the panel corresponding to the concatenation of ObsID 13251 and ObsID 12298 has a broken x-axis because there is a gap in spacecraft time between the two observations.

6 Application to X-ray Observations of FK Com

FK Com is an evolved yellow giant star (spectral type G4 III) with a mass of about 8.4 times that of the sun and at a distance of 550 ± 60 light-years. Typically, late-spectral-type giants are slow rotators, and thus do not maintain a strong magnetic dynamo that can sustain a high-temperature corona. However, FK Com is an extremely fast rotator, with equatorial velocities of 179 km/s, and a rotational period of 2.4 days (Korhonen *et al.*, 1999; Strassmeier, 2009). (For comparison, the Sun rotates at about 2 km/s at its equator.) Giant stars are formed when a sun-like star depletes the hydrogen at its core and thus transitions from producing energy though the fusion of hydrogen into helium to the fusion of heavier elements. Typical sun-like progenitor stars would require rotational speeds that would tear them apart in order to form a giant star with such a rapid rate of spin. Thus, the favored hypothesis for the formation of FK Com-like stars is that they are the result of a
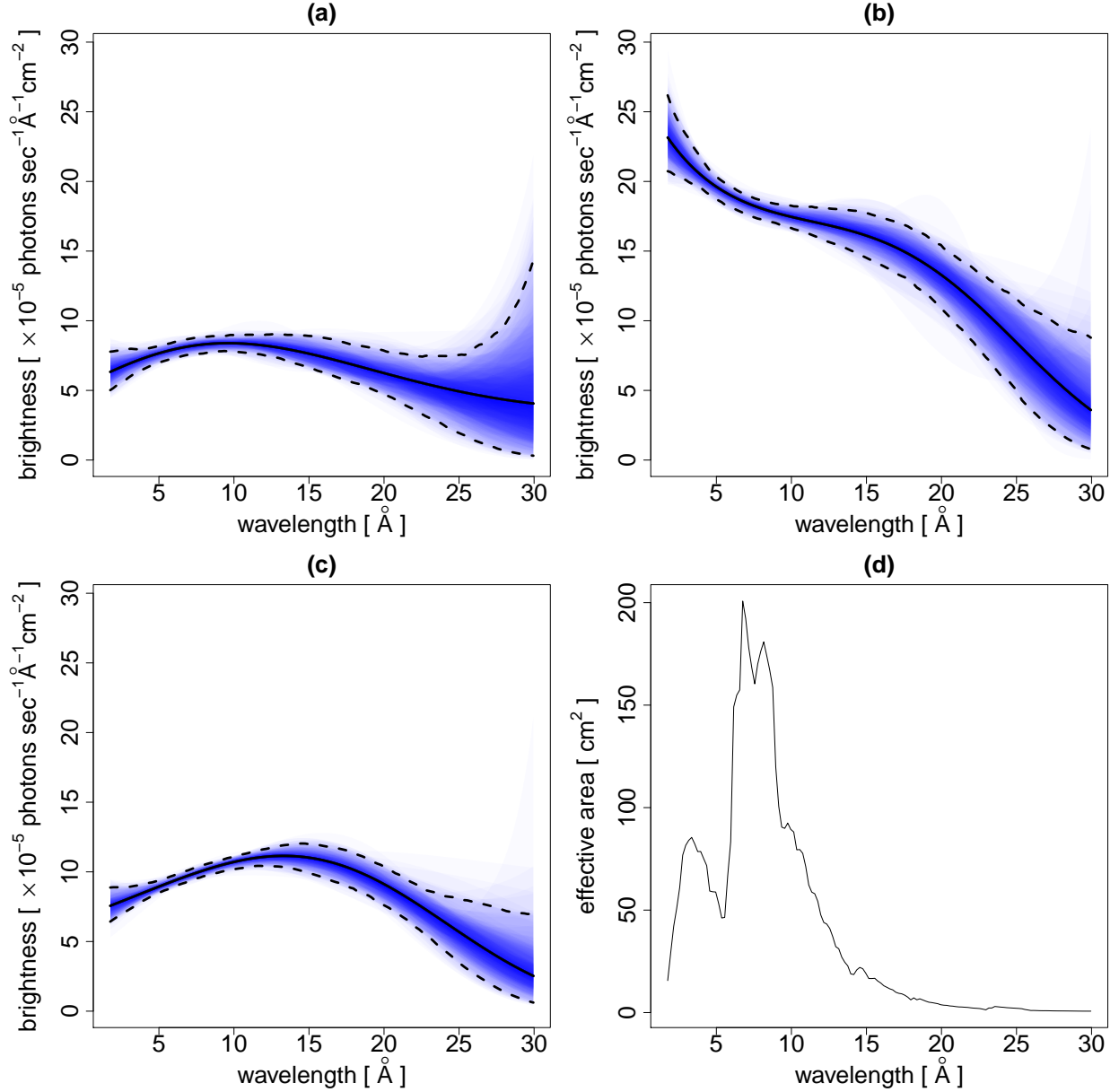
Figure 2: Panels (a) - (c): Overlay plots of the $\hat{f}$ fitted to the 200 simulations generated with the test function corresponding to ObsID 12299. Notice that all 200 simulations resulted in $\hat{B} = B$ and $\hat{\boldsymbol{\pi}} = \boldsymbol{\pi}$. The three panels represent the $B = 3$ homogeneous time intervals of ObsID 12299. The areas between the $\hat{f}$ computed with each simulated dataset and the underlying test function are colored and overlaid. Darker color represents more overlap of these areas and thus summarizes the concentration of the the fitted spectra around the underlying test function. The solid black line represents the smooth component, f, of test function 12299 and the dashed lines represent the 5% and 95% (pointwise) percentiles of $\hat{f}$. Similar plots for the other test functions appear in Appendix C. Panel (d): Plot of the effective area $\sum_{k=1}^{K} A_k(w)$ curve for test function 12299.

binary star system that has coalesced through a process of mass transfer, dramatically speeding up the stellar rotation in the process (Bopp and Stencel, 1981). If this is the case, the high rotation should induce a strong dynamo and consequently a dynamic corona, filled with high-temperature plasma and organized into loop-like structures via magnetic fields.

FK Com exhibits long duration spots—akin to sunspots—and flares associated with these spots. Optical Doppler imaging observations of FK Com demonstrate the existence of persistent spot-like features (Elstner and Korhonen, 2005) that remain localized to specific longitudes for long timescales (>months). Analysis of an X-ray flare observed with the XMM-*Newton* observatory showed that the flare is tied to these spots (Drake *et al.*, 2008). Numerical MHD modeling of the stellar wind and the magnetic structure (Cohen *et al.*, 2010) shows a highly complex coronal environment, with the most notable feature being even putatively stable coronal loops "wrapping around" due to the rotation. Thus, even though the corona appears to be solar-like in its origin, it is an extreme case of such coronae, and is therefore of considerable interest to both theoretical and observational studies.

A long-duration, high-resolution X-ray observation of FK Com was carried out with the *Chandra X-ray Observatory* over seven discontinuous time periods in April 2011. The data show that the source brightness is constantly fluctuating with occasional periods of singular flaring activity, see Figures 3 and 4. As a test case for our method, we apply it to these seven *Chandra* observations of FK Com. Each was preprocessed by discarding photons with wavelength less than 1.65 or greater than 31 Ångströms. The natural resolution of *Chandra* data of this sort is extremely high, resulting in a very large, but very sparse table of photon counts; typically there is at most one photon recorded in each wavelength-cross-time bin. We can obtain substantial computational savings with only a modest sacrifice of scientific information by reducing the resolution of the data. Massive events in the corona of FK Com, for example, require appreciable time and are associated with dramatic spectral shifts. Change points typically appear with frequencies on the order of tens of thousands of seconds, whereas the inherent bin width is about 3 seconds. In our analyses, we reduce this resolution and use temporal bins of width $\delta_t = 2000$ seconds. Similarly, the inherent wavelength bin width is 0.005 Ångströms and we set $\delta_w = 0.2$ Ångströms. Though a substantial reduction in spectral resolution, our analysis is much higher resolution than the standard hardness ratio analysis, which can be viewed as using two or three wavelength bins of width 10-15 Ångströms. Our fitted $\lambda(t_j, w_i)$ are plotted in Figure 5.

Table 3: Monte Carlo p-values for the test for change points in each of the seven observations of FK Com

| ObsID	12297	12356	13251	12298	13259	12357	12299
p-value	0.12	1.00	<0.01	0.01	0.14	1.00	<0.01

We also applied the Monte Carlo test for change points as described in Section 4.4; the resulting p-values for the seven datasets appear in Table 3. There is no evidence for change points in two of

19
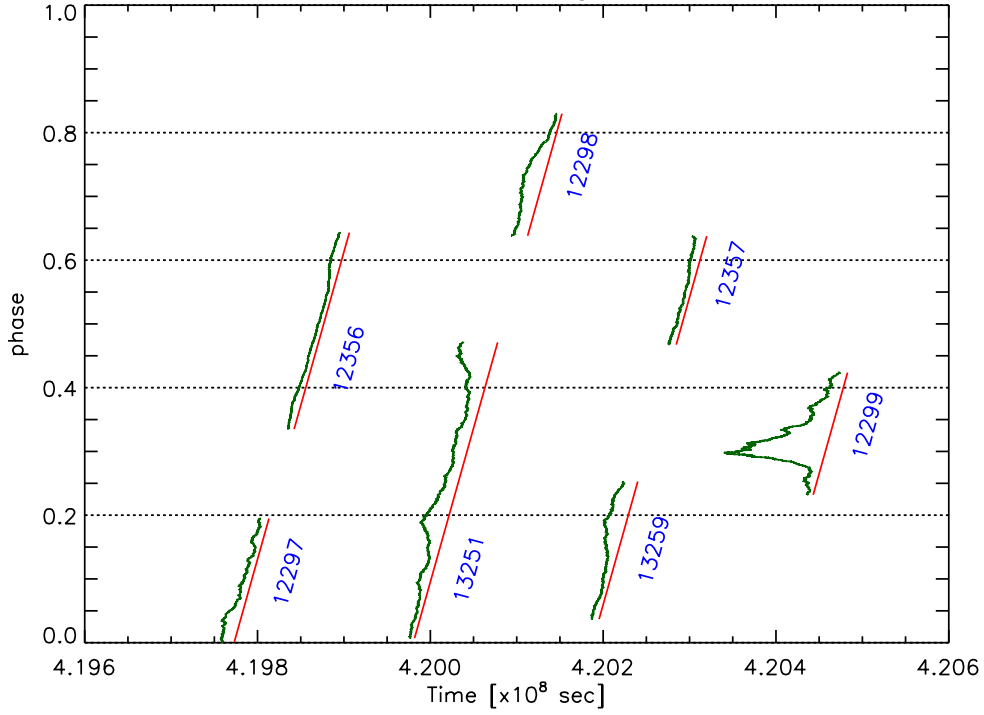
Figure 3: The phase-time light curve of FK Com. The source brightness (i.e., light curve) over the course of the *Chandra* observation is shown during periods when the source was observed. The horizontal axis represents spacecraft elapsed time, and the vertical axis represents the rotational phase of the star. The tilted red lines show the phase-time correspondence for each observation, showing the time and rotational phase coverage of the full dataset. Overlaid on the phase-time line, displayed as a deviation from it, are sparklines depicting the smoothed light curves (green curves). The segments are labeled by the ObsID numbers. This representation of the light curve shows that there is no obvious rotational signature in the light curve, and the variability is stochastic. Notice the large flare that occurs in ObsID 12299: our analysis shows that the spectrum varies significantly during this flare, see Figure 5.

the observations (ObsID 12356 and 12357), very weak evidence for two (ObsID 12297 and 13259), and strong evidence for the other three. This agrees with the fitted models illustrated in Figure 5, but quantifies the degree of evidence for the fitted change points. Finally, the two test functions that resulted in relatively low recovery rates of B in the simulations described in Section 5 correspond to the two data sets with the weakest evidence for the a change point, compare ObsID 12297 and 13259 in Table 2 (% correct $\hat{B}$) with Table 3 (p-value).

Although we find two or fewer change points in each observation, the number was not limited in our analyses. Operational constraints limit the continuous observation time available with the space-based *Chandra X-ray Observatory*. The change points that we observe stem from massive energetic shifts in the cornea of FK Com and take time to evolve. Keeping in mind that a single massive shift of this sort on the Sun would likely destroy human civilization, observing a total of seven in one (discontinuous) observation period spanning less than a fortnight (see Figure 3) is not

20
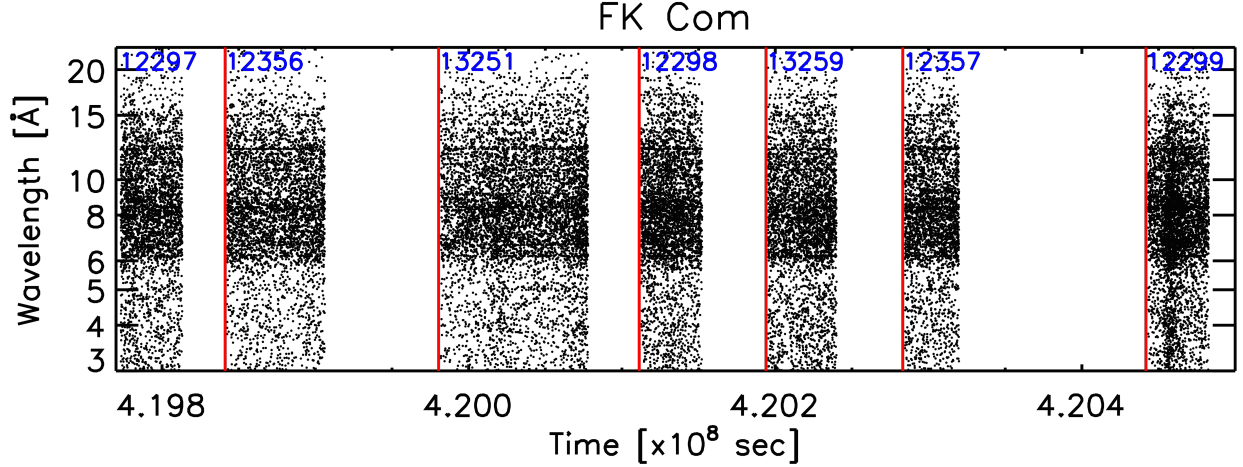
Figure 4: Observed data. Each point represents a photon detected with HETGS+ACIS-S configuration of *Chandra* (including both low- and high-energy gratings data). The wavelengths are given in Ångströms, and the time is in spacecraft clock seconds. The ObsIDs of each segment are marked to the right of the red vertical line representing the starting time of the observation segment. The collected data are sparse, but some changes in the density of points can be discerned even by eye, indicating temporal and spectral variations.

a small number.

This analysis indicates that large changes in intensities are accompanied by significant changes in the spectrum of the source. This is readily apparent in the results for ObsID 12299, which shows the observation separated into three segments with different intensities and spectral shapes; see ObsID 12299 in Figure 3 and row 7 of Figure 5. The spectrum changes to being dominated by high-energy photons due to an increase in high-temperature plasma when the flare is set off. As the flare decays, even when the observed intensity is relatively high, the spectrum can be seen returning to the pre-flare state, though with enhanced emission over the line-dominated region around 15 Ångströms.

From a theoretical point of view, this is not an obvious result, since for active stars like FK Com, it is believed that even the apparently quiescent corona emits X-rays as a superposition of a large number of weak flares (see, e.g., Kashyap *et al.*, 2002; Güdel, 2004). However, our analysis favors the idea that hot plasma could become trapped in stretched field lines wrapping around the star, and thus cool radiatively, and flares are only triggered sporadically when the magnetic strain becomes too large to be sustained. While a detailed modeling and interpretation of this picture is beyond the scope of this paper, we point out that even a simple application of a statistically principled method is enough to obtain an important result: The quiescent corona of FK Com is *unlike* the flaring corona, and in this sense is much more solar-like than previously anticipated.
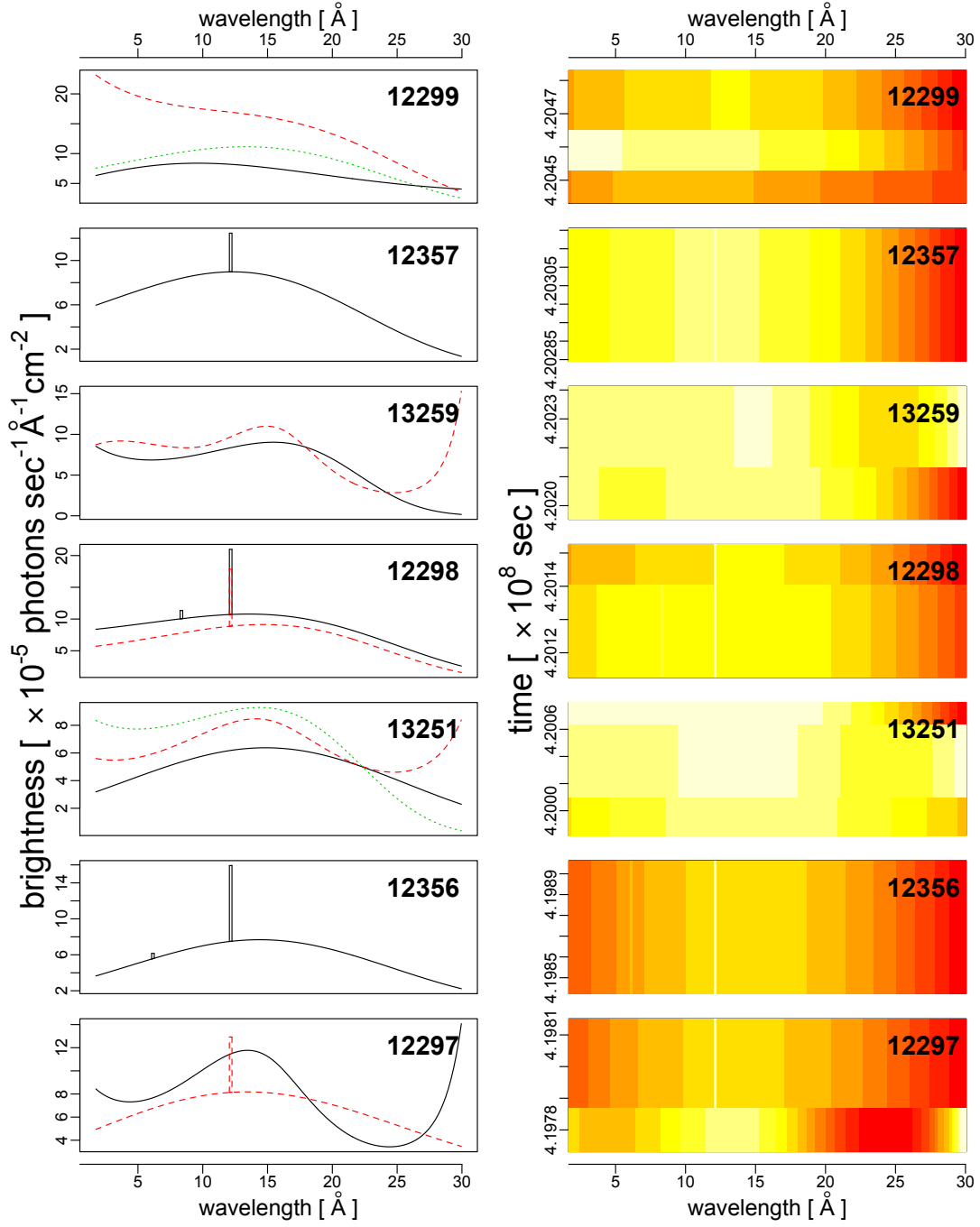
Figure 5: Fitted $\lambda(t_j, w_i)$ for seven *Chandra* observations of FK Com. The ObsID of each dataset appears in the title of each panel. The left plot in each row shows the fitted spectra for each identified time interval; the solid black, dashed red, and dotted green curves correspond to the first, second, and third time intervals, respectively. The right plot in each row is a heat maps that represents the best fit values of $\lambda(t_j, w_i)$. Time is given in elapsed spacecraft time.

22

7 Discussion

This article develops a novel approach for modeling change points in a class of marked Poisson processes, specifically the time-varying spectra of high-energy astrophysical sources. The approach includes a family of change point models, a practical algorithm for estimation, and an MDL criterion for selecting the best fitting model. Despite numerous reported applications where the MDL principle leads to reliable model selection criteria, its use in the statistical literature remains rather sparse. It is our hope that the present work, in which the MDL principle is applied to solve a complicated model selection problem that involves the "large p small n" scenario, will encourage and stimulate renewed research interests in this useful principle among statisticians.

Although the methods developed in this article are motivated by the search for change points in high-energy time-varying spectra, it is possible to extend them to include spatial variations as well. Typically, astrophysical images include sources at many different scales, ranging from unresolved point sources to complex structures that may span the field of view. Owing to instrumental point-spread functions, point sources in close proximity may overlap and in many cases are overlaid on larger scale structures. An extension that can account for change points in the spatial plane or in the spatial cross spectral hyperplane would be quite useful in astronomical data analysis. Statistically, this involves extending the dimension of the marks in the Poisson process from the one dimensional photon wavelengths to two dimension spatial locations or three dimensional spatial cross spectral measurements. This would, of course, involve extensions of the non-parametric models developed here for spectra to higher dimensional models for images or joint models for images and spectra.

A   Derivation of the MDL criterion (9)

This appendix derives (9), and it begins with the calculation of the term $\mathsf{CL}(\hat{\mathcal{M}})$ in (8). In Section 3, every candidate model $\mathcal{M}$ can be uniquely identified by a $\boldsymbol{\theta}$ and thus $\mathsf{CL}(\hat{\mathcal{M}}) = \mathsf{CL}(\hat{\boldsymbol{\theta}})$. Now to completely describe an $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^{\mathsf{T}}, \hat{\boldsymbol{\eta}}^{\mathsf{T}})^{\mathsf{T}}$, one could first specify which elements are nonzero, and then specify the actual values of these nonzero elements. As there are $\binom{P}{\|\hat{\boldsymbol{\beta}}\|_0}$ ways of arranging $\|\hat{\boldsymbol{\beta}}\|_0$ nonzero $\hat{\beta}_j$'s in P "slots", we need $\log_2 \binom{P}{\|\hat{\boldsymbol{\beta}}\|_0}$ bits to specify which $\hat{\beta}_j$'s are nonzero. Similarly we need $\log_2 \binom{N}{\|\hat{\boldsymbol{\eta}}\|_0}$ bits to specify which $\hat{\eta}_j$'s are nonzero.

Now for the actual values of the nonzero elements in $\hat{\boldsymbol{\theta}}$. As they are penalized maximum likelihood estimates, the approximate effective codelength for each of them is $\frac{1}{2}\log_2(NJ)$ (e.g., Rissanen, 1989), giving the total codelength for all nonzero elements as $\frac{1}{2}\|\hat{\boldsymbol{\theta}}\|_0 \log_2(NJ)$. Therefore

$$\mathsf{CL}(\hat{\mathcal{M}}) = \mathsf{CL}(\hat{\boldsymbol{\theta}}) = \frac{1}{2}\|\hat{\boldsymbol{\theta}}\|_0 \log_2(NJ) + \log_2 \binom{P}{\|\hat{\boldsymbol{\beta}}\|_0} + \log_2 \binom{N}{\|\hat{\boldsymbol{\eta}}\|_0}.$$

Since in practice $N \gg P$ so the second term is ignored, giving our final expression for $\mathsf{CL}(\hat{\mathcal{M}})$:

$$\mathsf{CL}(\hat{\mathcal{M}}) = \frac{1}{2}\|\hat{\boldsymbol{\theta}}\|_0 \log_2(NJ) + \log_2 \binom{N}{\|\hat{\boldsymbol{\eta}}\|_0}.$$

We remark that in many classical applications of two-part MDL where the number of possible parameters is small when compared to the number of data points, a penalty term similar to the last term in the above expression is typically ignored. However, for the present problem the number of potential parameters $P + N$ is not ignorable when compared to the number of data points NJ (where J could be small due to change points, see Section 4). Failing to consider this last term will lead to an overfitted model. This last term shares the same role as the additional penalty term in the Extended Bayesian Information Criterion (EBIC) proposed by Chen and Chen (2008). These authors have shown that the traditional BIC fails in the so-called "large p small n" scenario and an additional penalty term is required to guarantee consistency properties.

Lastly we need to calculate the term $\mathsf{CL}(\mathcal{D}|\hat{\mathcal{M}})$, which has been shown by Rissanen (1989) that it is given by the negative of the log likelihood. In our case this gives $\mathsf{CL}(\mathcal{D}|\hat{\mathcal{M}}) = -\sum_{i=1}^{N}\sum_{j=1}^{J} l_{\text{one}}(Y_{ij}; \hat{\boldsymbol{\theta}})$ with base 2 logarithm. Combining the expressions for $\mathsf{CL}(\hat{\mathcal{M}})$ and $\mathsf{CL}(\mathcal{D}|\hat{\mathcal{M}})$ and changing the base of the logarithm from 2 to e, we obtain (9).
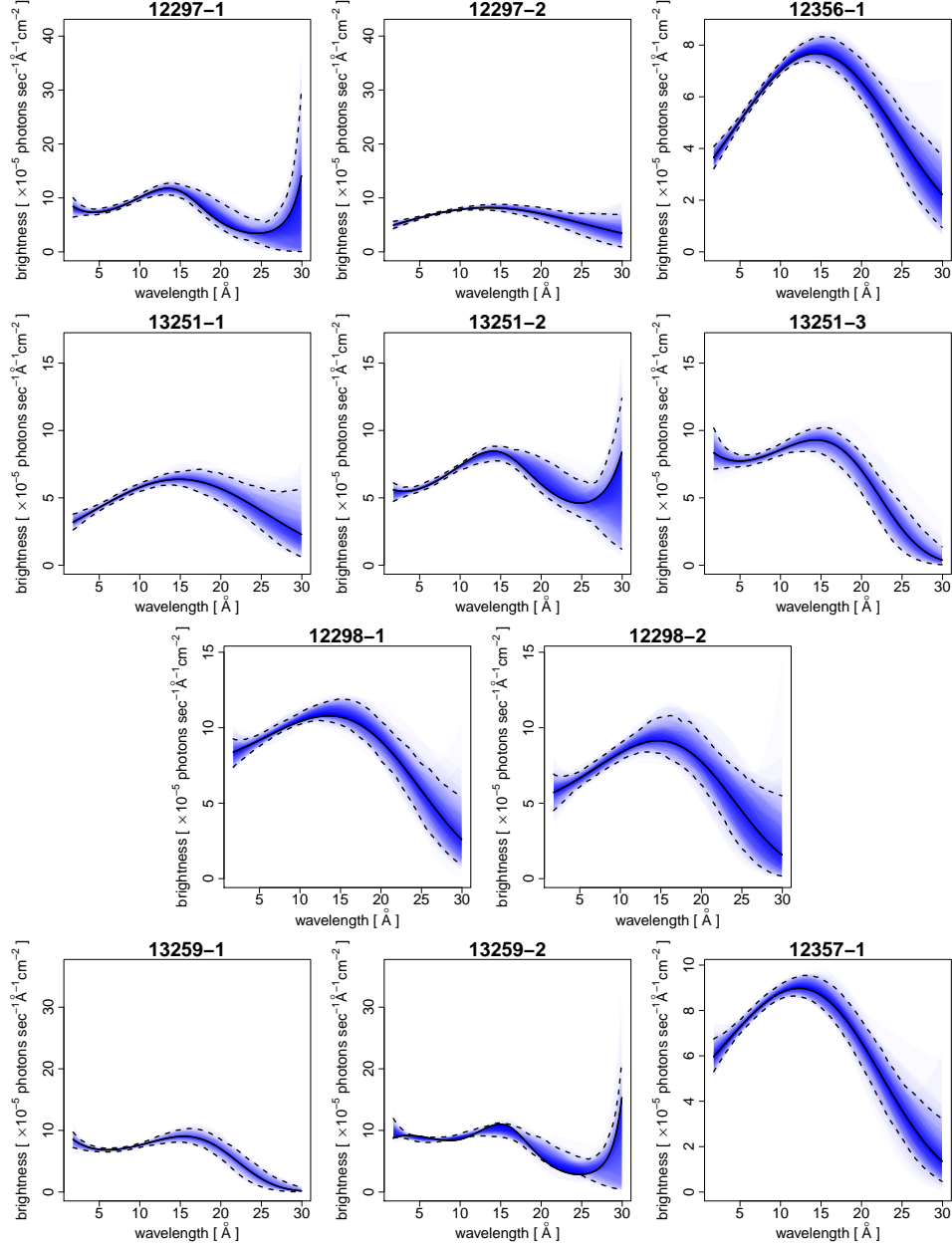
Figure 6: Similar to Figure 2 but for test functions 12297 (77%), 12356 (100%), 13251 (69%), 12298 (90%), 13259 (62%), and 12357 (100%). Results are only plotted for simulations with $\hat{B} = B$ and $\hat{\pi} = \pi$; the proportions of simulations for which $\hat{B} = B$ and $\hat{\pi} = \pi$ shown in the parentheses. Plots for the concatenated test function are not reported.

B Derivation of the MDL criterion (13)

This appendix derives the MDL criterion (13). As before, we need to calculate $\mathsf{CL}(\mathcal{D}) = \mathsf{CL}(\hat{\mathcal{M}}) + \mathsf{CL}(\mathcal{D}|\hat{\mathcal{M}})$. For model (10), the codelength $\mathsf{CL}(\hat{\mathcal{M}})$ for any candidate model $\hat{\mathcal{M}}$ can be decomposed

into
$$\mathsf{CL}(\hat{\mathcal{M}}) = \mathsf{CL}(B) + \mathsf{CL}(\boldsymbol{\pi}|B) + \mathsf{CL}(\hat{\boldsymbol{\Theta}}|B, \boldsymbol{\pi})$$

Since B is an integer, its codelength is $\mathsf{CL}(B) = \log_2 B$. Now for the codelength of $\boldsymbol{\pi}$. Recall that c_b is the length of the time segment b. Therefore knowing the values of $(c_1, \ldots, c_B)$ is equivalent to knowing the values of all change points, $(\pi_1, \ldots, \pi_{B-1})$. Thus it suffices to encode $(c_1, \ldots, c_B)$, and because they are integers, we have

$$\mathsf{CL}(\boldsymbol{\pi}|B) = \sum_{b=1}^{B} \log_2 p_b.$$

For the last term, we use the same arguments in Appendix A and obtain

$$\mathsf{CL}(\hat{\boldsymbol{\Theta}}|B, \boldsymbol{\pi}) = \sum_{b=1}^{B} \mathsf{CL}(\hat{\boldsymbol{\theta}}_b|B, \boldsymbol{\pi}) = \sum_{b=1}^{B} \left[\frac{1}{2}\|\hat{\boldsymbol{\theta}}_b\|_0 \log_2\{Nc(\boldsymbol{\pi})\} + \log_2 \binom{N}{\|\hat{\boldsymbol{\eta}}_b\|_0} \right].$$

Finally, $\mathsf{CL}(\mathcal{D}|\hat{\mathcal{M}})$ is given by the negative of the log likelihood (base 2) of the candidate model being considered, which gives

$$\mathsf{CL}(\mathcal{D}|\hat{\mathcal{M}}) = \sum_{i=1}^{N} \sum_{j=1}^{J} \sum_{b=1}^{B} I_b(t_j) l_{\text{one}}\{Y_{ij}; \hat{\boldsymbol{\theta}}_b(B, \boldsymbol{\pi})\},$$

with base 2 logarithm. Changing the base of the logarithm terms, we obtain (13).

C Supplement to Section 5

Overlay plots of the fitted $\hat{f}$ to the 200 simulations generated with the test functions corresponding to ObsID 12297, 12298, 12356,12357, 13251, and 13259 are shown in Figure 6.These test functions share the same effective area curve depicted in Figure 2(d).

References

Akman, V. and Raftery, A. (1986) Asymptotic inference for a change-point poisson process. *the Annals of Statistics*, 1583–1590.

Aue, A., Cheung, R. C. Y., Lee, T. C. M. and Zhong, M. (2014) Segmented model selection in quantile regression using the minimum description length principle. *Journal of the American Statistical Association*. To appear.

Aue, A. and Lee, T. C. M. (2011) On image segmentation using information theoretic criteria. *Annals of Statistics*, **39**, 2912–2935.

Bopp, B. and Stencel, R. (1981) The FK Comae stars. *The Astrophysical Journal*, **247**, L131–L134.

Buhmann, M. D. (2003) *Radial basis functions: theory and implementations.* Cambridge: Cambridge university press.

Carlin, B. P., Gelfand, A. E. and Smith, A. F. (1992) Hierarchical Bayesian analysis of changepoint problems. *Applied statistics*, 389–405.

Chan, H. P. and Zhang, N. R. (2007) Scan statistics with weighted observations. *Journal of the American Statistical Association*, **102**, 595–602.

Chen, J. and Chen, Z. (2008) Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, **95**, 759–771.

Chib, S. (1998) Estimation and comparison of multiple change-point models. *Journal of econometrics*, **86**, 221–241.

Cohen, O., Drake, J., Kashyap, V., Korhonen, H., Elstner, D. and Gombosi, T. (2010) Magnetic structure of rapidly rotating fk comae-type coronae. *The Astrophysical Journal*, **719**, 299–306.

Davis, R. A. and Yau, C. Y. (2013) Consistency of minimum description length model selection for piecewise stationary time series models. *Electronic Journal of Statistics*, **7**, 381–411.

Drake, J. J., Chung, S. M., Kashyap, V., Korhonen, H., Van Ballegooijen, A. and Elstner, D. (2008) X-ray spectroscopic signatures of the extended corona of fk comae. *The Astrophysical Journal*, **679**, 1522–1530.

Elstner, D. and Korhonen, H. (2005) Flip-flop phenomenon: observations and theory. *arXiv preprint astro-ph/0501343*, **326**, 278–282.

Friedman, J., Hastie, T. and Tibshirani, R. (2010) Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, **33**, 1.

Green, P. J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711–732.

Grünwald, P. D., Myung, I. J. and Pitt, M. A. (2005) *Advances in minimum description length: Theory and applications.* MIT press.

Güdel, M. (2004) X-ray astronomy of stellar coronae. *The Astronomy and Astrophysics Review*, **12**, 71–237.

Kashyap, V. L., Drake, J. J., Güdel, M. and Audard, M. (2002) Flare heating in stellar coronae. *The Astrophysical Journal*, **580**, 1118–1132.

Korhonen, H., Berdyugina, S., Hackman, T., Duemmler, R., Ilyin, I. and Tuominen, I. (1999) Study of fk comae berenices. i. surface images for 1994 and 1995. *Astronomy and Astrophysics*, **346**, 101–110.

Lai, T. L. and Xing, H. (2011) A simple Bayesian approach to multiple change-points. *Statistica Sinica*, **21**, 539.

Leclerc, Y. G. (1989) Constructing simple stable descriptions for image partitioning. *International Journal of Computer Vision*, **3**, 73–102.

Lee, H., Kashyap, V. L., van Dyk, D. A., Connors, A., Drake, J. J., Izem, R., Meng, X. L., Min, S., Park, T., Ratzlaff, P., Siemiginowska, A. and Zezas, A. (2011) Accounting for calibration uncertainties in X-ray analysis: Effective areas in spectral fitting. *The Astrophysical Journal*, **731**, 126–144.

Lee, T. C. M. (1997) *Some Models and Methods in Image Segmentation.* Ph.D. thesis, Macquarie University, Sydney, Australia.

Lee, T. C. M. (1998) Segmenting images corrupted by correlated noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**, 481–492.

Lee, T. C. M. (2002a) Automatic smoothing for discontinuous regression functions. *Statistica Sinica*, **12**, 823–842.

Lee, T. C. M. (2002b) On algorithms for ordinary least squares regression spline fitting: A comparative study. *Journal of Statistical Computation and Simulation*, **72**, 647–663.

Loader, C. R. (1992) A log-linear model for a poisson process change point. *The Annals of Statistics*, 1391–1411.

MacKay, D. J. C. (2003) *Information Theory, Inference, and Algorithms.* Cambridge, UK: Cambridge University Press.

McKee, M. (2012) Superflares' erupt on some Sun-like stars. *Nature News,*, May 16, 2012.

Mei, Y., Han, S. W. and Tsui, K.-L. (2011) Early detection of a change in poisson rate after accounting for population size effects. *Statistica Sinica*, **21**, 597.

Moreno, E., Casella, G. and Garcia-Ferrer, A. (2005) An objective Bayesian analysis of the change point problem. *Stochastic Environmental Research and Risk Assessment*, **19**, 191–204.

Ninomiya, Y. (2014) Change-point model selection via AIC. *Annals of the Institute of Statistical Mathematics*, 1–19.

Park, J. H. (2010) Structural change in US presidents' use of force. *American Journal of Political Science*, **54**, 766–782.

Park, T., Kashyap, V., Siemiginowska, A., van Dyk, D. A., Zezas, A., Heinke, C. and Wargelin, B. J. (2006) Bayesian estimation of hardness ratios: Modeling and computations. *The Astrophysical Journal*, **652**, 610–628.

Park, T., Krafty, R. T. and Sánchez, A. I. (2012) Bayesian semi-parametric analysis of Poisson change-point regression models: application to policy-making in Cali, Colombia. *Journal of applied statistics*, **39**, 2285–2298.

Raftery, A. and Akman, V. (1986) Bayesian analysis of a Poisson process with a change-point. *Biometrika*, 85–89.

Rissanen, J. (1989) *Stochastic Complexity in Statistical Inquiry*. World Scientific, Singapore.

Rissanen, J. (2007) *Information and Complexity in Statistical Modeling*. Springer.

Ruppert, D., Wand, M. P. and Carroll, R. J. (2003) *Semiparametric regression*. Cambridge: Cambridge university press.

Scargle, J. D. (1998) Studies in astronomical time series analysis. V. Bayesian blocks, a new method to analyze structure in photon counting data. *The Astrophysical Journal*, **504**, 405–418.

Scargle, J. D., Norris, J. P., Jackson, B. and Chiang, J. (2013) Studies in astronomical time series analysis. VI. Bayesian block representations. *The Astrophysical Journal*, **764**, 167–192.

Shen, J. J. and Zhang, N. R. (2012) Change-point model on nonhomogeneous Poisson processes with application in copy number profiling by next-generation DNA sequencing. *The Annals of Applied Statistics*, **6**, 476–496.

Strassmeier, K. G. (2009) Starspots. *The Astronomy and Astrophysics Review*, **17**, 251–308.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, **58**, 267–288.

van Dyk, D. A., Connors, A., Esch, D. N., Freeman, P., Kang, H., Karovska, M., Kashyap, V., Siemiginowska, A. and Zezas, A. (2006) Deconvolution in high-energy astrophysics: Science, instrumentation, and methods. *Bayesian Analysis*, **1**, 189–236.

van Dyk, D. A., Connors, A., Kashyap, V. and Siemiginowska, A. (2001) Analysis of energy spectra with low photon counts via Bayesian posterior simulation. *The Astrophysical Journal*, **548**, 224–243.

van Dyk, D. A. and Kang, H. (2004) Highly structured models for spectral analysis in high-energy astrophysics. *Statistical Science*, **19**, 275–293.

Worsley, K. J. (1986) Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika*, **73**, 91–104.

Xu, J., van Dyk, D. A., Kashyap, V. L., Siemiginowska, A., Connors, A., Drake, J. J., Meng, X. L., Ratzlaff, P. and Yu, Y. (2014) A fully Bayesian method for jointly fitting instrumental calibration and x-ray spectral models. *The Astrophysical Journal*, **794**, 97 (21pages).

Yao, Y.-C. (1988) Estimating the number of change-points via Schwarz' criterion. *Statistics and Probability Letters*, **6**, 181–189.

Yau, C.-Y., Tang, C.-M. and Lee, T. C. M. (2015) Estimation of multiple-regime threshold autoregressive models with structural breaks. *Journal of the American Statistical Association*, **110**, 1175–1186.

Zhang, N. R. and Siegmund, D. O. (2007) A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics*, **63**, 22–32.